Predicting rates of inbreeding: in populations undergoing selection

Predicting Rates of Inbreeding in Populations Undergoing Selection

John A. Woolliams* and Piter Bijma†

**Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, United Kingdom and †Animal Breeding and Genetics Group, Wageningen Institute of Animal Sciences, Wageningen Agricultural University, 6700 AH Wageningen, The Netherlands*

Manuscript received March 30, 1999
Accepted for publication December 6, 1999

ABSTRACT

Tractable forms of predicting rates of inbreeding (ΔF) in selected populations with general indices, nonrandom mating, and overlapping generations were developed, with the principal results assuming a period of equilibrium in the selection process. An existing theorem concerning the relationship between squared long-term genetic contributions and rates of inbreeding was extended to nonrandom mating and to overlapping generations. ΔF was shown to be $\sim\frac{1}{4}(1 - \omega)$ times the expected sum of squared lifetime contributions, where ω is the deviation from Hardy-Weinberg proportions. This relationship cannot be used for prediction since it is based upon observed quantities. Therefore, the relationship was further developed to express ΔF in terms of expected long-term contributions that are conditional on a set of selective advantages that relate the selection processes in two consecutive generations and are predictable quantities. With random mating, if selected family sizes are assumed to be independent Poisson variables then the expected long-term contribution could be substituted for the observed, providing $\frac{1}{4}$ (since $\omega = 0$) was increased to $\frac{1}{2}$. Established theory was used to provide a correction term to account for deviations from the Poisson assumptions. The equations were successfully applied, using simple linear models, to the problem of predicting ΔF with sib indices in discrete generations since previously published solutions had proved complex.

WRAY and Thompson (1990) proved a fundamental relationship between the sum of squared long-term genetic contributions of ancestors and rates of inbreeding for random mating populations in discrete generations. One consequence of this relationship was that rates of inbreeding were tied to the numerator relationship matrix for the first time. This narrowed the conceptual gap between the central parameter for genetic evaluation of individuals using best linear unbiased prediction and one of the key properties of a breeding scheme. Another important consequence was to set out in a formal way a model for the mechanics of inheritance of selected advantage, a concept that Robertson (1961) had introduced but had left unclarified. An achievement of the methods of Wray and Thompson (1990) was to obtain, for the first time, accurate predictions of ΔF in mass selection through modeling pathway extensions. However, this was done by using a recursive algorithm, so that although the mechanics were clear, the overall structure of the prediction remained obscure.

Woolliams *et al.* (1993) advanced the understanding of the structure of the prediction by obtaining a closed form for the prediction of ΔF. It was shown to have terms involving variances of family size in one generation, with additional terms for the proliferation or reduction of ancestral lines over many generations that could be predicted as a result of the selective advantage of the ancestor. Furthermore, it was clear that under equilibrium conditions, the model would lend itself to geometric summation of terms across generations. This led to simple forms for the expected long-term contribution of an ancestor. Wray *et al.* (1994) extended the methods to index selection, although the form of the model is a hybrid of the approach of Woolliams *et al.* (1993) and Hill (1972), since the conditional arguments of pathway extension that had been carried out for mass selection were found to be too complex for index selection. Nevertheless, worthwhile predictions were made available in a tractable form.

Santiago and Caballero (1995) used an approach that made no direct reference to the theory of contributions to predict ΔF in mass selection. They obtained a neater closed form for ΔF than that derived by Woolliams *et al.* (1993) through an argument based on total drift, relating the change through selection to loss of genetic variance. Unlike the previous work of Wray and Thompson (1990) and Woolliams *et al.* (1993), who had considered the population in relation to an unselected base generation, Santiago and Caballero (1995) developed predictions based upon equilibrium genetic variance. Nomura (1996) extended the approach of Santiago and Caballero (1995) to mass selection with overlapping generations but with the important restriction that the males and females selected from a cohort remain the same in both number and identity throughout the breeding life of the cohort.

This article examines the issues raised by the work

Corresponding author: J. A. Woolliams, Roslin Institute (Edinburgh), Roslin, Midlothian EH25 9PS, United Kingdom.
E-mail: john.woolliams@bbsrc.ac.uk

TABLE 1

Notation used to derive Equations 1–27

t,u	Time variables
n_c, q	Number of breeding categories, indexed by q
m, f	Denotes the two sexes in discrete generations, *i.e.*, q = m (male) or f (female)
r_i, $r_{i(q)}$	Observed long-term genetic contribution of individual i (in category q)
$r_{i,u}(q, t)$	The genetic contribution of individual i born at time u to selected parents of sex q born at time t
F_t, ΔF	Inbreeding coefficient at time t, and rate of inbreeding
ω	Deviation from Hardy-Weinberg equilibrium
X_q, X	Number of parents in category q and a simple monoecious population, respectively
$C_u(t)$	Sum of squared contributions for individuals born at time u to selected parents at time t
C	Converged sum of squared contributions, independent of time in an equilibrium
L	Generation interval
$\boldsymbol{s}_{i(q)}$	Set of selective advantages for individual i in category q
$\mu_{i(q)}$	Expected contribution of i in category q conditional upon $\boldsymbol{s}_{i(q)}$
$\sigma^2_{i(q)}$	Variance of contribution of i in category q conditional upon $\boldsymbol{s}_{i(q)}$
n_i	Number of selected offspring of i
$\theta_{n,i}$	Expected number of selected offspring of i conditional upon $\boldsymbol{s}_{i(q)}$
$V_{n,i}$	Variance of the number of selected offspring of i, conditional upon $\boldsymbol{s}_{i(q)}$
$V_{n,\mathrm{dev},i}$	Deviation of $V_{n,i}$ from Poisson, *i.e.*, $V_{n,\mathrm{dev},i} = V_{n,i} - \theta_{n,i}$
α_q, $\boldsymbol{\beta}_q$	Linear model for $\mu_{i(q)} = \alpha_q + \boldsymbol{\beta}_q^{\mathrm{T}}(\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}}_q)$

described above. First, the relationship between ΔF and the realized long-term genetic contributions is extended to include nonrandom mating and overlapping generations. Second, an important result for the prediction of ΔF is developed by demonstrating a relationship between ΔF and the expected squared long-term contribution conditional on the selective advantages for random mating. Finally, as an example of application, predictions of ΔF for sib indices, previously considered by Wray *et al.* (1994), are reexamined using the equilibrium methods for expected long-term contributions developed by Woolliams *et al.* (1999) and compared to results from simulation.

RELATIONSHIP BETWEEN ΔF AND LONG-TERM GENETIC CONTRIBUTIONS

This section discusses the relationship between ΔF and realized long-term genetic contributions. In doing so, it derives the expected increase in homozygosity at the level of a neutral locus in contrast to the matrix method of Wray and Thompson (1990). The notation that is used is shown in Table 1. The model for the population is assumed, for the present, to have discrete generations with X_{m} male parents and X_{f} female parents. For calculation of inbreeding coefficients every allele is considered as unique in the base population ($t = 0$). It does not matter if the base generation has the structure of an unselected and unrelated population.

Discrete generations: Consider one of these alleles in the base population at a neutral locus (say allele B). Let the gene frequency at time t, in the parents of sex q that have been selected to produce generation $t + 1$, be denoted by $P_B(q, t)$. The gene frequency can be described in terms of genetic contributions similar to Equation 1 of Woolliams *et al.* (1999). Let A_i be the gene frequency of an allele B in individual i, where $A_i = 1$, $\frac{1}{2}$, or 0 if i is BB, $B\cdot$, or $\cdot\cdot$, respectively (where $\cdot$ represents any other allele), then the individual gene frequencies can be treated as breeding values for frequency. The average of the gene frequency in the parents of sex q in generation t is given by

$$P_B(q, t) = \sum_i r_{i,0}(q, t)A_{i,0} + \sum_{u=1}^{t}\sum_i r_{i,u}(q, t)a_{i,u}, \quad (1)$$

where $r_{i,u}(q, t)$ is the genetic contribution of individual i born at time u to the parents of sex q at time t, with breeding value for frequency of allele B given by $A_{i,u}$ and Mendelian sampling terms $a_{i,u} = A_{i,u} - \frac{1}{2}(A_{\mathrm{sire}} + A_{\mathrm{dam}})$. Equation 1 separates out the base generation, which provides the foundation alleles, and subsequent generations, which influence the frequency of the allele through the Mendelian sampling of their parent alleles. The variance of the Mendelian sampling terms will depend on A_{sire} and A_{dam}; $\mathrm{Var}(a_{i,u}) = 0$ if both A_{sire} and A_{dam} are homozygotes, $\frac{1}{8}$ if they are both heterozygotes, or $\frac{1}{16}$ otherwise. Since B is unique, $A_{i,0}$ is 0 for all individuals except for one individual for which $A_{i,0} = \frac{1}{2}$. The genetic contribution of an individual to the generation of its birth is $r_{i,t}(m, t) = X_{\mathrm{m}}^{-1}$ if i is male or 0 if i is female, and $r_{i,t}(f, t) = X_{\mathrm{f}}^{-1}$ if i is female or 0 if i is male.

Initially assume that there is random mating. For any generation the probability of homozygotes for B is obtained from the product of the gene frequencies in the male and in female parents and is $P_B(m, t)P_B(f, t)$. The inbreeding coefficient F_t for the neutral locus is then the sum over all distinct alleles at the locus,

$$F_t = \sum_{\mathrm{alleles}}\sum_i r_{i,0}(m, t-1)r_{i,0}(f, t-1)A_{i,0}^2 + \sum_{\mathrm{alleles}}\sum_{u=1}^{t-1}\sum_i r_{i,u}(m, t-1)r_{i,u}(f, t-1)a_{i,u}^2, \quad (2)$$

where $r_{i,u}(q, t-1)$ is the average contribution to parents of sex q at time $t - 1$. (Note the breeding values and Mendelian sampling terms will depend on the allele but this dependence has not been made explicit to spare notation.) For each allele the cross-product terms in $A_{i,0}$ $A_{j,0}$ are zero since $A_{i,0} = 0$ except for a single individual. Since the Mendelian sampling terms from different individuals are independent of all other terms for a neu-

tral locus, all cross-products of the Mendelian sampling terms are zero.

More precisely, for each allele and each ancestor, the term $\Sigma_i r_{i,u}(m, t-1) r_{i,u}(f, t-1) a_{i,u}^2$ should be the sum of products of contributions of the ancestor to each male and female mating pair:

$$\sum_i \sum_{\text{mates } (j(m), j^*(f))} r_{i,u}(j(m), t-1) r_{i,u}(j^*(f), t-1) a_{i,u}^2. \quad (3)$$

This will account for any nonrandom mating of parents. For a neutral locus, the covariance between r_i and a_i will be 0 (Woolliams and Thompson 1994; Woolliams *et al.* 1999), and the expectation of Equation 3 is $E[\Sigma_i \Sigma_{\text{mates } (j(m), j^*(f))} r_{i,u}(j(m), t-1) r_{i,u}(j^*(f), t-1)] E[a_{i,u}^2]$. Let the first of these, the expectation of the cross-products of contributions to mates, be $C_u(t-1)$. Note that (i) $C_{t-1}(t-1) = 0$ since an individual without offspring cannot contribute to both sexes and (ii) the first term in Equation 2 is $\frac{1}{2}C_0(t-1)$ since $A_{i,0}^2$ has a value $\frac{1}{4}$ for each of its two alleles and 0 otherwise.

Assume equilibrium values for (i) the deviation from Hardy-Weinberg frequencies arising from the nonrandom mating (ω, equivalent to α_1 of Caballero and Hill 1992) and (ii) ΔF, attained by generation 2 (this assumption is removed later); then Equation 2 can be further simplified using results given in appendix a, namely, $\Sigma_{\text{alleles}} E[a_{i,u}^2] = \frac{1}{4}$ for $u = 1$ and $\frac{1}{4}(1-\omega)(1-\Delta F)^{u-1}$ for $u \geq 2$. Therefore,

$$E[F_{t+1}] = \tfrac{1}{2}C_0(t) + \tfrac{1}{4}C_1(t) + \sum_{u=2}^{t-1} \tfrac{1}{4}(1-\omega)(1-\Delta F)^{u-1} C_u(t) \quad (4)$$

$$E[F_t] = \tfrac{1}{2}C_0(t-1) + \tfrac{1}{4}C_1(t-1) + \sum_{u=2}^{t-2} \tfrac{1}{4}(1-\omega)(1-\Delta F)^{u-1} C_u(t-1). \quad (5)$$

Subtracting (5) from (4) and rearranging terms,

$$E[F_{t+1} - F_t] = \tfrac{1}{2}C_0(t) - \tfrac{1}{2}C_0(t-1) + \tfrac{1}{4}C_1(t) - \tfrac{1}{4}C_1(t-1) + \tfrac{1}{4}(1-\omega)(1-\Delta F)C_2(t) + \sum_{u=2}^{t-2} \tfrac{1}{4}(1-\omega)(1-\Delta F)^{u-1} \times [(1-\Delta F)C_{u+1}(t) - C_u(t-1)]. \quad (6)$$

Assuming equilibrium, then a steady state of pedigree development will occur and the expectation of the cross-products will be determined by the number of generations over which they have developed, *i.e.*, $C_u(t) = C_{u-1}(t-1)$ since both terms represent contributions $t-u$ generations after the birth of the ancestor. This is not a strong assumption in the context of the problem since in the absence of an equilibrium there would be no single ΔF to predict.

Therefore, the terms in $C_u(t)$ can be modified to terms in $C_{u-1}(t-1)$, and each term of the sum within the square brackets of Equation 6 can be reduced to $-\Delta F C_u(t-1)$. After repeating this process for the $C_2(t)$ term [and temporarily neglecting the term in $\omega\Delta F C_1(t-1)$],

$$E[F_{t+1} - F_t] = \tfrac{1}{2}C_0(t) - \tfrac{1}{2}C_0(t-1) + \tfrac{1}{4}C_1(t) - \tfrac{1}{4}\omega C_1(t-1) - \Delta F\,[\tfrac{1}{4}C_1(t-1) + \sum_{u=2}^{t-2} \tfrac{1}{4}(1-\omega)(1-\Delta F)^{u-1} \times C_u(t-1)]. \quad (7)$$

For large enough t, the terms in $C_u(t)$ will converge for a given u. Therefore, $\frac{1}{2}C_0(t) \approx \frac{1}{2}C_0(t-1)$, and $\frac{1}{4}C_1(t) - \frac{1}{4}\omega C_1(t-1) \approx \frac{1}{4}(1-\omega)C_1(t-1)$; then by adding and subtracting the term $\frac{1}{2}\Delta F C_0(t)$,

$$E[F_{t+1} - F_t] = \tfrac{1}{2}\Delta F C_0(t-1) + \tfrac{1}{4}(1-\omega)C_1(t-1) - \Delta F[\tfrac{1}{2}C_0(t-1) + \tfrac{1}{4}C_1(t-1) + \sum_{u=2}^{t-2} \tfrac{1}{4}(1-\omega)(1-\Delta F)^{u-1} \times C_u(t-1)]. \quad (8)$$

Finally, note $E[F_{t+1} - F_t] = \Delta F E[1 - F_t]$ and that the term in square brackets in Equation 8 is $E[F_t]$, giving

$$\Delta F = \tfrac{1}{4}(1-\omega)C_1(t-1)[1 - \tfrac{1}{2}C_0(t)]^{-1}. \quad (9)$$

This result holds for t large enough for contributions from early generations to have converged. If it is assumed that the base generation used for defining the inbreeding coefficients was chosen to be part of a period of equilibrium, then $C_1(t-1) = C_0(t) = C$,

$$\Delta F = \tfrac{1}{4}(1-\omega)C[1 - \tfrac{1}{2}C]^{-1} \approx \tfrac{1}{4}(1-\omega)C, \quad (10)$$

where C is the sum of squared converged contributions for a generation, chosen arbitrarily within the period of equilibrium. Including the term neglected between Equations 6 and 7 would replace $[1 - \frac{1}{2}C]^{-1}$ by $[1 - (\frac{1}{2} + \frac{1}{4}\omega)C]^{-1}$. For random mating, omitting the term $[1 - \frac{1}{2}C]^{-1}$ leads to an underestimate with a fractional error of $\sim\frac{1}{2}C$, which in turn is $\sim 2\Delta F$.

Since $C = E[\Sigma_i \Sigma_{\text{mates } (j(m), j^*(f))} r_{i,u}(j(m), t)\, r_{i,u}(j^*(f), t)]$ for large $u \ll t$, for any i the terms $r_{i,u}(j(m), t)$ and $r_{i,u}(j(f), t)$ converge to the same value for all j in generation t providing the population mixes. This value will be the long-term contribution of ancestor i to the population, denoted by r_i. This will occur with or without random mating. Thus $C = E[\Sigma_i r_i^2]$ for a generation of ancestors, which leads to

$$\Delta F \approx \tfrac{1}{4}(1-\omega)\, E\left[\sum_i r_i^2\right] \quad (11)$$

$$E[\Delta F] \approx \tfrac{1}{4}(1-\omega)(X_m E[r_{i(m)}^2] + X_f E[r_{i(f)}^2]). \quad (12)$$

In Equation 12, the expectations are conditional on the individual i being a selected ancestor; however, since

$r_i = 0$ for an unselected ancestor, Equation 12 can also be given as

$$E[\Delta F] \approx \tfrac{1}{4}(1 - \omega)(T_m E[r^2_{i(m)}] + T_f E[r^2_{i(f)}]), \quad (13)$$

where T_m and T_f are the number of candidates for selection in each sex and the expectation is for a candidate (*i.e.*, it is not conditional on i being selected). ($E[\Delta F]$ is used in Equations 12 and 13, rather than simply ΔF, to emphasize that the result is an expectation over replicate populations.)

This result was obtained for $\omega = 0$ by Wray and Thompson (1990) but the derivation differs in several aspects. First, in the derivation of Wray and Thompson the base was unselected and therefore not in equilibrium at the start of the selection process, and this led to an impression that the contributions used for estimating rates of inbreeding must be the generation after an unselected base. It is now evident that the choice of generation on which the estimate is obtained is arbitrary except that it is at the start of some period of local equilibrium during which some "equilibrium ΔF" may exist. Second, the derivation using the probability of homozygosity for an assumed allele is of value since the proof of Wray and Thompson (1990) is heavily based upon the properties of the numerator relationship matrix. Third, it extends the result to incorporate nonrandom mating, although the result was given without proof by Woolliams and Thompson (1994). Caballero and Hill (1992) noted that the result of Wray and Thompson (1990) was a poor predictor of ΔF with nonrandom mating and it is now clear why this was so.

Even though the development of the pedigree may be in equilibrium (which will imply the genetic variance being selected upon is in equilibrium) this does not imply that equilibrium values of ω and ΔF for the alleles defined in the arbitrary base are immediately attained. Equation 4, using appendix a, assumes that these parameters were in equilibrium for the Mendelian sampling in generation 2. However, the following argument shows that this does not affect the result. Assume the equilibrium conditions have not been attained by generation 2; then for this generation plus a small number of generations following (*i.e.*, up to attainment of equilibrium) there will be terms of the form $\delta C_u(t)$ in Equation 4 and $\delta C_u(t - 1)$ in Equation 5. Providing t is sufficiently large compared to the period of attainment, these terms will cancel in Equation 6 since $C_u(t)$ is a convergent series. Thus Equations 10–13 will hold for the equilibrium values of ω and ΔF.

Overlapping generations: If ΔF is taken per unit time then the structure of the preceding proof holds. The reduction in the variance of the Mendelian sampling term over initial cohorts, before an equilibrium ΔF/unit time is established, is not straightforward since it will depend upon the age structure of the population; but the previous argument used to overcome deviations from equilibrium can be applied. However, one distinction in overlapping generations is that the base generation will contain the equivalent of L cohorts, where L is the period of time over which the long-term contributions sum to one, since this is the period required for the population to turn over a generation for those genes destined to remain in the population in the long-term. Woolliams *et al.* (1999) show this genetic generation interval is different from the average age of the parents when there are selection advantages between groups (see also Bijma and Woolliams 1999). To balance (8) there is a need to add and subtract terms of magnitude $\frac{1}{2}C_0(t)$ (ΔF/generation) or equivalently $\frac{1}{2}C_0(t)L$ (ΔF/unit time), where L is the generation interval. Thus the error term in Equation 10 is $[1 - \frac{1}{2}CL]^{-1}$, and consequently ignoring this term results in an underestimate with a fractional error of $2 \times$ (ΔF per generation). Equation 11 is obtained by summing over all individuals born in a single cohort. With overlapping generations, individual ancestors within cohorts will have different life histories, since they will be used at different breeding ages or for different purposes. If X_q is the number of individuals with a lifetime breeding profile categorized by q, then the approximation will be

$$E[\Delta F/\text{unit time}] \approx \tfrac{1}{4}(1 - \omega) \sum_{\text{categories } q} X_q E[r^2_{i(q)}], \quad (14)$$

where the expectations are over the squared contributions from a single cohort and are conditional on selection in category q. Although the approach is different, Equation 14 is equivalent to the result of Hill (1972, 1979) when random selection and random mating is assumed. However, Equation 14 clearly shows that the rate of inbreeding is related to the sum of squared lifetime contributions irrespective of selection and nonrandom mating.

RELATIONSHIP BETWEEN ΔF AND EXPECTED CONTRIBUTIONS

Since ΔF is proportional to $E[r_i^2]$, the task of predicting ΔF in selected populations would be made easier if tractable and general methods for calculating expected squared contributions were available. However, $E[r_i^2] = \mu_i^2 + \sigma_i^2$ and consequently there is a need to predict both the mean and variance of the contributions. Commonly, the prediction of means is a simpler task than the prediction of variances. General methods for predicting expected long-term contributions in selected populations have been developed by Woolliams *et al.* (1999). The objective of the following section is to obtain a relationship between the variance of the long-term contributions and their expectations, which will then permit development of general methods for the prediction of $E[r_i^2]$ and consequently for ΔF. The relationship will need to assume random mating and is developed by conditioning on the selective advantage(s), $\boldsymbol{s}_i$, for an ancestor. The selective advantage(s) of the ancestor, if

inherited, will partly determine the breeding success of its descendants, with diminishing impact over generations. The proof uses the result $E[r_i^2] = E_s[r_2^2|\boldsymbol{s}_i] = E_s[\mu_i^2] + E_s[\sigma_i^2]$, where $\mu_i = E[r_i|\boldsymbol{s}_i]$ and $\sigma_i^2 = \mathrm{Var}[r_i|\boldsymbol{s}_i]$, and the subscript s on the E indicates that the expectation is being taken over the selective advantages.

Monoecious population: The proof is simplest in the case of a monoecious diploid population of X parents in discrete generations without selfing. Random mating is assumed ($\omega = 0$). Extension to overlapping generations and to two sexes follows by analogy but is complicated by the need for matrices, and so this extension is made in appendix b. The long-term contribution of individual i is given by

$$r_i = \tfrac{1}{2} \sum_{\text{offspring } j} r_j. \tag{15}$$

These sums may be restricted to the selected offspring since unselected offspring have no long-term contribution. It is assumed that conditional on the selective advantage $\boldsymbol{s}_i$ of the parent i, the genetic contribution of the offspring is independent of the number of offspring selected from parent i (denote this number by n_i). Then from Equation 15,

$$E[r_i|\boldsymbol{s}_i,n_i] = \tfrac{1}{2}n_iE[r_j|\boldsymbol{s}_i, j \text{ offspring of } i] \tag{16}$$

$$\mathrm{Var}[r_i|\boldsymbol{s}_i,n_i] = \tfrac{1}{4}n_i\mathrm{Var}[r_j|\boldsymbol{s}_i, j \text{ offspring of } i]. \tag{17}$$

Equation 17 requires random mating. Let $\theta_{n,i}$ and $V_{n,i}$ be the mean and variance of $n_i|\boldsymbol{s}_i$; then

$$\mu_i = \tfrac{1}{2}\theta_{n,i}E[r_j|\boldsymbol{s}_i, j \text{ offspring of } i]. \tag{18}$$

The derivation of μ_i in a general genetic framework was described by Woolliams *et al.* (1999).

The variance σ_i^2 is derived using the statistical result that the unconditional variance is the expectation of the conditional variance plus the variance of the conditional expectation. Applying this result to Equations 16 and 17 gives

$$\sigma_i^2 = \tfrac{1}{4}\theta_{n,i}\mathrm{Var}[r_j|\boldsymbol{s}_i, j \text{ offspring of } i] + \tfrac{1}{4}V_{n,i}E[r_j|\boldsymbol{s}_i, j \text{ offspring of } i]^2. \tag{19}$$

Assume now that the number selected from parent i has a Poisson distribution. For example, this would be the case if litter size before selection had a Poisson distribution. Then $\theta_{n,i}$ can replace $V_{n,i}$ in the second term of Equation 19 to obtain

$$\sigma_i^2 = \tfrac{1}{4}\theta_{n,i}(\mathrm{Var}[r_j|\boldsymbol{s}_i, j \text{ offspring of } i] + E[r_j|\boldsymbol{s}_i, j \text{ offspring of } i]^2), \tag{20}$$

which can be recognized as

$$\sigma_i^2 = \tfrac{1}{4}\theta_{n,i}\, E[r_j^2|\boldsymbol{s}_i, j \text{ offspring of } i]. \tag{21}$$

If expectations are now taken over $\boldsymbol{s}_i$, Woolliams and Bijma (1999) show that by assuming an equilibrium there is no covariance between $\theta_{n,i}$ and $E[r_j^2|\boldsymbol{s}_i, j \text{ offspring of } i]$. A heuristic explanation is that if there were a covariance, then this would result in selection for increased squared contributions, breaking the assumption of equilibrium. The right-hand side is then equal to $\tfrac{1}{2}E_s[r_i^2|\boldsymbol{s}_i]$, since $E_s[\theta_{n,i}] = 2$. Therefore,

$$E_s[\sigma_i^2] = \tfrac{1}{2}E_s[r_i^2|\boldsymbol{s}_i] = \tfrac{1}{2}E_s[\mu_i^2] + \tfrac{1}{2}E_s[\sigma_i^2], \tag{22}$$

which leads to the result that

$$E_s[\sigma_i^2] = E_s[\mu_i^2]. \tag{23}$$

Finally, if X is the number of parents in each generation, then

$$E[\Delta F] \approx \tfrac{1}{4}X\, E_s[r_i^2] = \tfrac{1}{4}X(E_s[\mu_i^2] + E_s[\sigma_i^2]) = \tfrac{1}{2}XE_s[\mu_i^2]. \tag{24}$$

The power of this result is that it requires only the mean conditional on the selective advantages to be modeled, which can be done for a wide class of genetic structures using the methods of Woolliams *et al.* (1999). Note that the set of selective advantages used for conditioning must completely describe the interrelationship between one generation of selection and the next. This is embodied in the assumption that conditioning on the selective advantage $\boldsymbol{s}_i$ removes associations between the number of offspring selected and the subsequent success of the offspring. For example, the mates of the individual provide a selective advantage that must be accounted for (Woolliams and Thompson 1994; Santiago and Caballero 1995).

One of the critical assumptions of the proof leading to (24) is that the selected family sizes are distributed as a Poisson variable. However, departures from this will occur, for example, (i) when the litter sizes are not Poisson; (ii) when negative covariances between full-sibs and between half-sibs are induced by using sib indices for selection; (iii) when selection intensity becomes large; and (iv) when there are common environmental variances associated with litters. (The occurrence of the last two causes will depend on the model chosen for $\boldsymbol{s}$, which is addressed in the discussion.)

To account for this deviation let $V_{n,i} = \theta_{n,i} + V_{n,\mathrm{dev},i}$ in Equation 19, where $V_{n,\mathrm{dev},i}$ may be positive or negative according to the circumstances. Then the component in $\theta_{n,i}$ can be treated as previously and Equation 21 becomes

$$\sigma_i^2 = \tfrac{1}{4}\theta_{n,i}E[r_j^2|\boldsymbol{s}_i, j \text{ offspring } i] + \tfrac{1}{4}V_{n,\mathrm{dev},i}E[r_j|\boldsymbol{s}_i, j \text{ offspring of } i]^2 \tag{25}$$

and Equation 23 becomes

$$E_s[\sigma_i^2] = E_s[\mu_i^2] + \tfrac{1}{2}E_s[V_{n,\mathrm{dev},i}E[r_j|\boldsymbol{s}_i, j \text{ offspring of } i]^2] \tag{26}$$

with the result

$$\Delta F = \tfrac{1}{2}XE_s[\mu_i^2] + \tfrac{1}{8}XE_s[V_{n,\mathrm{dev},i}E[r_j|\boldsymbol{s}_i, j \text{ offspring of } i]^2]. \tag{27}$$

Anticipating an observed result, the magnitude of terms involving $\boldsymbol{s}_i$ in $E[r_j|\boldsymbol{s}_i, j$ offspring of $i]$ contributes very little to the second term of Equation 27 and only the constant term, independent of $\boldsymbol{s}_i$, needs be considered. In the current context $E[r_j|\boldsymbol{s}_s, j$ offspring of $i] \approx X^{-1}$ and the second term in Equation 27 becomes $\frac{1}{8}E_s[V_{n,\mathrm{dev},i}]/X$. For example, in mass selection with fixed litter sizes, Santiago and Caballero (1995) used the approximation that $E_s[V_{n,\mathrm{dev},i}] \approx -n_o^{-1}$, where n_o is the number of offspring per parent, with the result that the correction for the deviation from Poisson is $(-8T)^{-1}$ where T is the total number of individuals born.

One of the benefits of Equation 24 is that the rate of inbreeding can be obtained from predicting means, often using regression techniques. Accounting for deviations from the Poisson distribution introduces the need for estimating variances of family size to obtain Equation 27. Nevertheless, the multigenerational problem of estimating the variance of a long-term genetic contribution has been reduced to estimating the variance of family size after selection in a single generation.

Extension to overlapping generations: With overlapping generations, individuals within a cohort that are selected to breed at any point in their lifetime can be divided into breeding categories. These categories are defined by the age of breeding, how often, and for what purpose the individual breeds. Categories are particularly important in selection. As an example, consider mass selection where all selected individuals can have progeny born at ages 1, 2, or 3. If the population is making genetic progress the average merit of individuals born 3 years ago is less than the average merit of an individual born 1 year ago. Therefore an offspring of a 3-year-old parent will have a selective disadvantage compared to an offspring of a 1-year-old parent and so is expected to make a smaller genetic contribution in the long-term (see Bijma and Woolliams 1999). If an individual is a parent at all ages then its genetic contribution is expected to be greater than an individual chosen for breeding only at a single age. Breeding purpose is also important: if one group of parents are given more mating opportunities, then these would be expected to have more offspring and, other factors being equal, ultimately a greater long-term genetic contribution.

For these reasons partition of the selected individuals into categories is necessary to obtain the general result. It is assumed that the categories are defined so that an individual belongs to a single category that describes its lifetime genetic contribution. To continue the example of mass selection, where the only distinction among parents is the breeding age, there would be potentially seven categories. If $\{x\}$ denotes age x at breeding, then these categories are {1}, {2}, {3}, {1, 2}, {1, 3}, {2, 3}, {1, 2, 3}. The number of categories will inevitably depend on the complexity of the breeding scheme, but the essential point is that they can be defined and enumerated. Let n_c be the number of categories indexed from $q = 1 \ldots n_c$, and $\mu_{i(q)}$ be the expected long-term contribution of individual i in category q conditional on its selective advantage $\boldsymbol{s}_{i(q)}$ with variance $\sigma^2_{i(q)}$. The steps given above in Equations 16–27 for a single category remain the same but changes are needed since terms must be redefined as vectors and matrices. The notation to develop the argument therefore becomes more complex but the result remains simple. For this reason the proof is given in appendix b. The conclusion is that if family sizes after selection are assumed to be distributed as independent Poisson variables, then

$$E[\Delta F] = \tfrac{1}{2}\sum_q X_q E_s[\mu^2_{i(q)}]. \qquad (28)$$

This simple result shows that the rate of inbreeding, when approximated by the sum of squared contributions, is equal to one half of the sum of the squares of expected lifetime contributions. Instead of using the observed contribution, as in Equation 12, the expected contribution can be substituted, but this is done at the cost of changing the coefficient from $\frac{1}{4}$ to $\frac{1}{2}$. This is because the expected contribution is being used to model both the mean and the variance.

As previously, for a parent from category q, define the matrix $\boldsymbol{V}_{n(q),\mathrm{dev}}$ of size $n_c \times n_c$ to be the (co)variance matrix for the number of selected offspring in each of the n_c categories, expressed as deviations from independent Poisson variances. For each q, neglecting terms in $\boldsymbol{s}$ (for empirical reasons given earlier), there will be a term δ_q defined by $\boldsymbol{\alpha}^T\boldsymbol{V}_{n(q),\mathrm{dev}}\boldsymbol{\alpha}$, where α is the vector with the qth element equal to the expected long-term contribution for an individual from category q, *i.e.*, $E_s[\mu_{i(q)}] = \alpha_q$. Note δ_q may be negative since it is a variance deviation and is not a variance. This term is introduced in Equation B6 of appendix b. From appendix b we arrive at

$$E[\Delta F] = \tfrac{1}{2}\sum_q X_q E_s[\mu^2_{i(q)}] + \tfrac{1}{8}\sum_q X_q \delta_q. \qquad (29)$$

Although the proof has been based upon a monoecious diploid organism with no selfing, the extension to a dioecious organism is clear from the proof for overlapping generations. Having discrete generations with two sexes is identical to having two categories, *i.e.*, males and females. Finally note that, other than assuming an equilibrium and random mating, there have been no assumptions on the type of selection index used, the nature of the genetic variation, or the population structure.

APPLICATIONS AND RESULTS

Sib indices in discrete generations: The theory is illustrated by selection on a general sib index of the form $I = b_1(P - \overline{P}_{\mathrm{fs}}) + b_2(\overline{P}_{\mathrm{fs}} - \overline{P}_{\mathrm{hs}}) + b_3\overline{P}_{\mathrm{hs}}$, where P is the phenotype of the candidate, $\overline{P}_{\mathrm{fs}}$ is the phenotypic mean of its full-sibs (including candidate), and $\overline{P}_{\mathrm{hs}}$ is the phenotypic mean of its half-sibs (including candidate and

TABLE 2

Genetic parameters for a population selected with a sib index

X_m, X_f, d	Number of male and female parents and mating ratio $d = X_f/X_m$
n_m, n_f, n_o	Number of male and female offspring in a full-sib family, $n_o = n_m + n_f$
P, $\bar{P}_{fs}$, $\bar{P}_{hs}$	Phenotype of candidate and its full- and half-sib family means
I, b_1, b_2, b_3	Index and weights for selection $I = b_1(P - \bar{P}_{fs}) + b_2(\bar{P}_{fs} - \bar{P}_{hs}) + b_3\bar{P}_{hs}$
p_m, p_f	Selection proportions for males and females: $p_m = (n_m d)^{-1}$, $p_f = n_f^{-1}$
v_m, v_f, i_m, i_f, k_m, k_f	Truncation points, intensities of selection, and variance reduction coefficients
σ_I^2, σ_A^2, ρ_I	Variance of the index, total genetic variance, and accuracy of selection
σ_{Am}^2, σ_{Af}^2, σ_e^2	Genetic variance among selected sires and dams and residual variance
$A_{i(hs)}$, $A_{j(fs)}$	Mean breeding value of the half-sib family of sire i and the mean breeding value of the full-sib family of dam j expressed as a deviation from the half-sib family
$\nu(A_{i(hs)})$	Variance of $A_{i(hs)}$ and similarly defined for $A_{i(fs)}$: $\nu(A_{i(hs)}) = \sigma_{Am}^2 + \sigma_{Af}^2/d$, $\nu(A_{i(fs)}) = \sigma_{Af}^2(1 - d^{-1})$
ρ_{FS}, ρ_{HS}	Correlation of indices among full-sibs and half-sibs, respectively: $\rho_{HS} = [b_3^2\nu(A_{i(hs)}) - b_2^2\nu(A_{i(fs)})(d-1)^{-1}]/\sigma_I^2$; $\rho_{FS} = [b_3\nu(A_{i(hs)}) + b_2\nu(A_{i(fs)}) - b_1^2(\frac{1}{2}h_0^2 + \sigma_e^2)n_o^{-1}]/\sigma_I^2$

In the initial unselected base population, the phenotypic variance was assumed to be 1 and the initial heritability h_0^2.

full-sibs). Mass selection is a special case, with $b_1 = b_2 = b_3 = 1$ (or any constant >0). This formulation was used also by Wray *et al.* (1994) in their study of rates of inbreeding. Every generation, the highest ranking X_m sires and X_f dams are selected as parents for the next generation. Each sire is mated at random to $d = X_f/X_m$ dams and each dam produces a total of n_o offspring, n_m male, and n_f female, which are available for selection in the next generation. The unselected base population is assumed to have a phenotypic variance of 1 with a heritability of h_0^2 for the selected trait. Additional notation used for the sib index is shown in Table 2. An example is given at each step and this is a selection scheme for $X_m = 20$, $X_f = 60$, $n_m = n_f = 4$, with index weights $b_1 = 1$, $b_2 = 1.5$, and $b_3 = 2$. The principal parameters for this scheme are presented in Table 3 for easy reference.

In Wray *et al.* (1994) the selective advantages were based on the breeding values $A_{i(x)}$, and this approach is adopted here but slightly modified. A sire i has one selective advantage, namely, its own breeding value plus the average breeding value of its d mates (*i.e.*, its mate group) and this aggregate value is denoted by $A_{i(hs)}$. A dam i has two selective advantages: first, the selective advantage of its mate ($A_{i(hs)}$) and second, its own breeding value expressed as a deviation from the average breeding value of the mate group to which it belongs (denoted $A_{i(fs)}$). The average breeding value of the full-sib family from dam i is $\frac{1}{2}(A_{i(hs)} + A_{i(fs)})$. Thus, in this hierarchical scheme, $\boldsymbol{s}_{i(m)} = (A_{i(hs)})$, and $\boldsymbol{s}_{i(f)} = (A_{i(hs)}, A_{i(fs)})^T$. The two selective advantages for a dam are independent.

Expected long-term genetic contributions were modeled following Woolliams *et al.* (1999) as $E[r_{i(q)}|\boldsymbol{s}_{i(q)}] = \mu_{i(q)} = \alpha_q + \boldsymbol{\beta}_q^T(\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}}_q)$, where $\boldsymbol{s}_{i(q)}$ denotes the vector of selective advantages for a selected individual of sex q expressed as a deviation from the mean of its contemporaries $\bar{\boldsymbol{s}}_q$, $\boldsymbol{\beta}_q$ is the vector of regression coefficients of $r_{i(q)}$ on $\boldsymbol{s}_{i(q)} - \bar{\boldsymbol{s}}_q$, α_q is the mean contribution of selected parents of sex q, and ^{T} denotes the transpose. In the parameterization used, the mean of $A_{i(fs)}$ is always zero. To simplify the notation it is assumed that $A_{i(hs)}$ is already expressed as a deviation from the mean of the contemporary group, and so $\bar{\boldsymbol{s}}_q$ is omitted from this point onwards.

Step 1. Prediction of expected contributions: The prediction of expected genetic contributions is covered in detail by Woolliams *et al.* (1999). The current article only summarizes the procedure for a sib index, without derivation. Prediction of $\mu_{i(q)}$ requires the prediction of $\boldsymbol{\alpha} = (\alpha_{\mathbf{m}}, \alpha_{\mathbf{f}})^T$ and $\boldsymbol{\beta} = (\beta_m^T, \beta_f^T)$. In discrete generations, $(\alpha_{\mathbf{m}}, \alpha_{\mathbf{f}}) = [1/(2X_m), 1/(2X_f)]$ always. Solutions for $\boldsymbol{\beta}$ are obtained applying the method of Woolliams *et al.* (1999), using Bulmer's (1980) equilibrium genetic variances. A summary of equations used is given in appendix c. For the example $(\alpha_m, \alpha_f) = (0.0250, 0.0083)$, $\boldsymbol{\beta} = (0.0447, 0.0149, 0.0130)$.

Step 2. Rates of inbreeding assuming Poisson variances: From step 1, $\mu_{i(m)} = [0.0250 + 0.0447A_{i(hs)}]$. The expected squared mean is a simple sum of squared terms: $X_m E[\mu_{i(m)}^2] = X_m[0.0250^2 + 0.0447^2\nu(A_{i(hs)})(1 - X_m^{-1})]$. The $(1 - X_m^{-1})$ term accounts for variances about the sample mean of the selected group rather than the true mean.

The terms arising from $X_f E[\mu_{i(f)}^2]$ are calculated analogously. Since the two selected advantages of the females

TABLE 3

Principal parameters, as described in Table 2, for the example selection scheme used throughout

$X_m = 20$	$X_f = 60$	$d = 3$	$n_m = n_f = 4$	$h_0^2 = 0.4$	$(b_1, b_2, b_3) = (1.0, 1.5, 2.0)$
$p_m = 0.083$	$p_f = 0.25$	$v_m = 1.383$	$v_f = 0.674$	$i_m = 1.839$	$i_f = 1.271$
$k_m = 0.838$	$k_f = 0.759$	$\sigma_I^2 = 1.331$	$\sigma_A^2 = 0.302$	$\sigma_{Am}^2 = 0.050$	$\sigma_{Af}^2 = 0.052$
$\rho_I = 0.636$	$\rho_{FS} = 0.390$	$\rho_{HS} = 0.205$	$\nu(A_{i(hs)}) = 0.269$		$\nu(A_{i(fs)}) = 0.140$

are mutually independent, the expected mean squared is simply the sum of squared terms. The expected long-term contribution of a female parent is

$$\mu_{i(f)}\ [0.0083 + 0.0149A_{i(\mathrm{hs})} + 0.0130A_{i(\mathrm{fs})}]$$

and the sum of squared means is

$$X_f E[\mu^2_{i(f)}] = X_f[0.0083^2 + 0.0149^2\nu(A_{i(\mathrm{hs})})(1 - X_m^{-1}) + 0.0130^2\nu(A_{i(\mathrm{fs})})].$$

As previously mentioned, the term is defined as a deviation from the mean over all ancestors so $\nu(A_{i(\mathrm{fs})})$ requires no correction.

The rate of inbreeding ignoring deviations from Poisson variances is predicted from $\Delta F = \frac{1}{2}(X_m E[\mu^2_{i(m)}] + X_f E[\mu^2_{i(f)}]) = (0.0227 + 0.0090)/2 = 0.0158$.

Step 3. Correction for deviations of V_n from Poisson variances: Deviations from Poisson variances can be accounted for by correcting the rate of inbreeding using Equation 28, where $\delta_q = \boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{V}_{n(q),\mathrm{dev}}\boldsymbol{\alpha}$ and $\mathbf{V}_{n(q),\mathrm{dev}}$ is the (2×2) matrix with (co)variances of the number of selected offspring of a parent of sex q (q = m, f) as a deviation from independent Poisson variances. The calculation of the deviation from Poisson family variance for fixed numbers of selection candidates per full-sib family is described in appendix d. The approach adopted was derived in detail by Burrows (1984), although extension to two sexes was required and the method was made more flexible by incorporating results from Mendell and Elston (1974). Applying the method to the example gives

$$\boldsymbol{V}_{n(m),\mathrm{dev}} = (0.186, 0.751 | 0.751, -0.079),$$

$$\boldsymbol{V}_{n(f),\mathrm{dev}} = (0.020, 0.159 | 0.159, -0.154).$$

The total correction to the predicted ΔF is 0.0016, and the prediction, using Equation 29, is 0.0175. The mean ΔF derived from 1000 simulations was 0.0183 (SE = 0.0001).

General fit: Extensive simulations were carried out assuming an infinitesimal model with factorial combinations of X_m = 20, 40, 80; d = 1, 2, 3 (and 5 for X_m = 20, 40); total offspring of 4, 8, and 16 per full-sib family equally divided between sexes; and with h^2 = 0.1, 0.2, 0.4, and 0.6; weights used were (1.0, 0.75, 0.5) for $d >$ 1 [changed to (1.0, 0.75, 0.75) for d = 1] and (1.0, 1.5, 2.0) for $d > 1$ [changed to (1.0, 1.5, 1.5) for d = 1]. Classical weights were also examined since these weights were the subject of the study of Wray *et al.* (1994), although they are suboptimal after the first round of selection from an unselected base population. Results have been tabulated and summarized by Woolliams and Bijma (1999).

With weights (1.0, 0.75, 0.5, or 0.75) the accuracy was excellent for all schemes, with all errors <4%. With weights (1.0, 1.5, 1.5, or 2.0) accuracy was also very good, accurately tracking trends with the changes in the parameters and with a large majority of errors <2% with the exception of d = 3, h^2 = 0.4, where underestimates of up to 8% were observed. The trends in rates of inbreeding were also accurately tracked with classical weights with no increases in the magnitude of the errors, even though schemes had rates of inbreeding >0.03.

The most serious trend in the errors was a pattern of underprediction characterized by high mating ratio and large family sizes (both of which increase the selection intensity) and increased family weights. More surprisingly, the errors also increased with the numbers of parents at a constant d (*i.e.*, X_m = 20, X_f = 60 compared to X_m = 80, X_f = 240), and also the errors were not present for h^2 = 0.01 and increased sharply as h^2 increased. To explore these errors further, the long-term contributions for selected males were plotted against $A_{i(\mathrm{hs})}$ for the following schemes with d = 3, weights (1.0, 1.5, 2.0): I, X_m = 20, h^2 = 0.4, n_o = 16; II, X_m = 80, h^2 = 0.4, n_o = 16; III, X_m = 80, h^2 = 0.01, n_o = 16; and IV, X_m = 80, h^2 = 0.4, with n_o = 4. The results for simulated (S) and predicted (P) were as follows: I, S = 0.0231, P = 0.0220; II, S = 0.0070, P = 0.0058; III, S = 0.0028, P = 0.0029; IV, S = 0.0037, P = 0.0037. Note that scheme II is simply scheme I with four times the number of parents and expected long-term contributions of I are consequently four times bigger than II. The prediction of ΔF for scheme II is close to (but not precisely) $\frac{1}{4}$ of that for I. However, the ratio of the simulated ΔF for scheme II compared to I was closer to $\frac{1}{3}$, *i.e.*, much greater than would be expected from scaling. Serious prediction error occurs only for scheme II.

Figure 1 shows that the accuracy of prediction with low h^2 (scheme III) is because the linear model used is a good fit (*i.e.*, the contributions are a simple linear regression on the selective advantage) and similarly for low selection intensity (scheme IV). However, for both the other two schemes the linear model predicts a substantial proportion of the selected males to have negative contributions, although rates of inbreeding are accurately predicted in one case (scheme I) but not in the other (scheme II).

Closer replicate-by-replicate analysis shows that despite the expectation, the substantially greater variance of contributions (approximately proportional to $\Delta F/X_m$) in scheme I obscures the nonlinearity in the majority of replicates. When both linear and quadratic terms for the selective advantage were included in a regression model for observed contributions, the quadratic term was not statistically significant (defined here as $P <$ 0.01) in >60% of the replicates. In contrast, for scheme II, this percentage was <15%. Thus the accuracy of prediction depends on the goodness-of-fit of the linear model within a replicate, so more parents may promote greater proportional prediction errors, even though these errors will be associated with lower rates of inbreeding.

The pattern of the correction for deviations from Poisson distribution for selected family sizes is worth

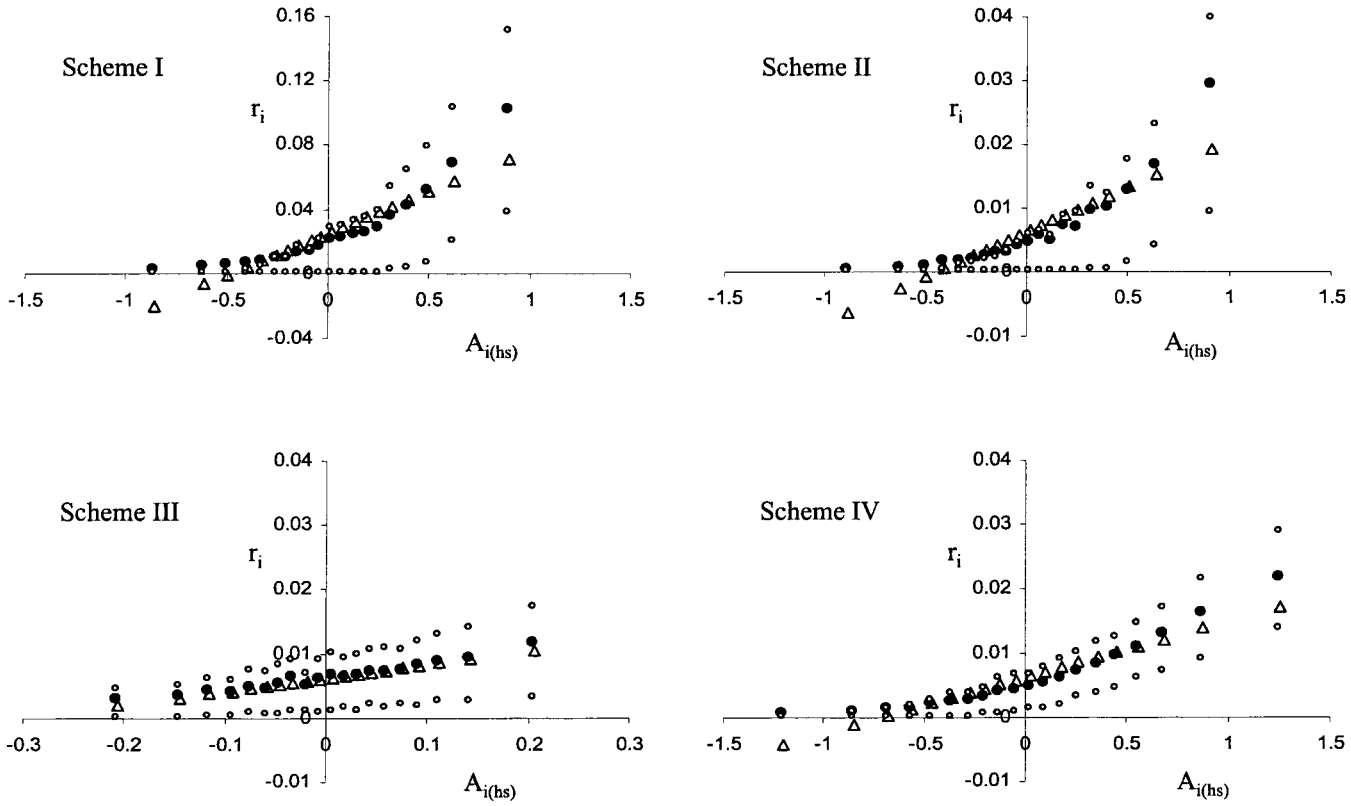

Figure 1.—The expected long-term contribution and lower and upper quartiles obtained from simulation (as a function of the selective advantage $A_{i(hs)}$), together with the expected long-term contribution predicted from assuming a linear model for four example schemes. The curves obtained from simulation are the result of sampling 8000 individuals. The following schemes all have $d = 3$ with weights (1.0, 1.5, 2.0): I, $X_m = 20$, $h^2 = 0.4$, $n_o = 16$; II, $X_m = 80$, $h^2 = 0.4$, $n_o = 16$; III, $X_m = 80$, $h^2 = 0.01$, $n_o = 16$; and IV, $X_m = 80$, $h^2 = 0.4$, $n_o = 4$. △, linear prediction; ●, simulated expectation; ○, lower and upper quartiles.

noting. These corrections are negative for b_2, $b_3 < 1$, reduce in size as the index weights increase, and were generally positive for b_2, $b_3 > 1$. For mass selection, $b_1 = b_2 = b_3 = 1$, the correction is of the order of $-1/(8T)$.

DISCUSSION

The theory described in this article provides a powerful tool for predicting rates of inbreeding in selected populations and for providing insights into the forces that contribute to the rate of loss of variation. The relationship of Wray and Thompson (1990) has been derived directly from consideration of identity by descent and has been extended to cover overlapping generations and nonrandom mating. Applicability was then advanced by showing how expected long-term contributions, which are predictable by general methods, can be used in place of observed long-term contributions to predict the rates of inbreeding, if random mating was assumed. Finally, the methods were applied to sib indices in discrete generations, for which the previous solutions were complex (Wray *et al.* 1994). In doing so, some insight was gained into the origin of the prediction errors, and these appeared to arise from the goodness-of-fit of the models used to implement the theory rather than those used to derive it.

Theory: The first theorem relating the rate of inbreeding in a population to the squared long-term contributions was previously derived by Wray and Thompson (1990) but the proof here has several useful extensions. In contrast to Wray and Thompson (1990), the proof is direct in using identity by descent rather than properties of the numerator relationship matrix, and it also incorporates nonrandom mating and overlapping generations. The simplest relationship ($\Delta F \approx \frac{1}{4}\Sigma r_i^2$) is not exact and was shown to underestimate the rate of inbreeding by a fraction of the order of ($2\Delta F$), providing there was no major deviation from random mating, and is therefore small for any practical scheme. In overlapping generations, with rates of inbreeding per unit time and per generation both of interest, it is shown that this error is $2(\Delta F/\text{generation})$, where the generation interval was defined by the period over which the long-term genetic contributions sum to 1.

The importance of the relationship between rates of inbreeding and squared genetic contributions is that it holds for selected populations, with no assumptions on the form of selection, providing (i) the genes are ultimately mixed, and (ii) an equilibrium exists over which

a stable ΔF may be defined. A further caveat is that the rate obtained applies to a neutral, unlinked gene. The extension of other relationships to predict ΔF in selected populations does not always hold. For example, using the relationship $\mathrm{Var}(\delta q) = q(1 - q)\Delta F$, where q is the frequency of a neutral gene and δq is the change in frequency per unit time, will not hold if selection is not random since it assumes mutual independence of δq over consecutive intervals. The increments, δq, are also correlated for overlapping generations due to the many intervals over which the progeny of a single parent may be selected. As a consequence the justification for the proof by Hill (1979) for ΔF with overlapping generations is invalid, even in the absence of genetic selection, although the result is correct and agrees with the previous proof of Hill (1972). Closer examination of Hill (1979) shows that its justification lies in an intuitive argument for the relationship that was to be proved later by Wray and Thompson (1990). Consequently the methods derived here may be seen to arise as a natural development of the results of Hill (1972, 1979) for selected populations.

The form of Equation 4 shows that the sum of squared long-term contributions for any given cohort may be usefully interpreted in the absence of an equilibrium. The sum of squared contributions for a cohort is the proportion of the new variation (the Mendelian sampling variance) arising from within that cohort that is lost to the population in the long term. This includes all mutational variance arising in prior generations, since the choice of base is arbitrary. Therefore the sum of squared contributions of cohorts (particularly those still to converge!) is important, irrespective of equilibrium, and provides a meaningful measure of risk, and merits attention in both breeding and conservation schemes. The operational tools described by Grundy *et al.* (1998) are based upon controlling sums of squared contributions of cohorts and have meaning and validity beyond the infinitesimal model (*e.g.*, Villanueva *et al.* 1999). However, there are clearly greater problems in providing deterministic predictive tools to analyze population dynamics if the assumption of equilibrium is removed, and those provided by Woolliams *et al.* (1999) assume this equilibrium.

The second, novel theorem derived in this article is concerned with showing how the formulas with observed long-term contributions may be translated into formulas with expected long-term contributions. The latter are advantageous since they use predictable entities. The major change is that the expected can be substituted for the observed, providing the constant of proportionality is increased from $\frac{1}{4}$ to $\frac{1}{2}$. The critical step in the proof is that the error variance of a long-term contribution given the selective advantage is related to the square of its mean, *i.e.*, the coefficient of variation is relatively constant. Apart from random mating, the scope of this proof is very broad and is applicable to overlapping generations. The validity of the derivation was checked using general sib-indices as an example in discrete generations, and a companion article (Bijma *et al.* 2000) provides verification in overlapping generations with mass selection with lifetime selection, thereby removing a serious restriction of Nomura (1996). The limitation to random mating arises from Equation 17, although in one special case, partial full-sib mating with no selection, the analysis can be completed (using results of Ghai 1965) and shown to agree with the results of Caballero and Hill (1992). This provides an indirect verification of Equation 13 for nonrandom mating.

Woolliams *et al.* (1999) show how the expected long-term contribution may be calculated in general for different inheritance models (*e.g.*, imprinted variation, maternal additive, or sex-linked variation) with different selection indices (sib indices or best linear unbiased predictors). Using long-term contributions follows the path of Wray and Thompson (1990) and Woolliams *et al.* (1993) and differs from Santiago and Caballero (1995; mass selection in discrete generations) and Nomura (1996; a special case of mass selection with overlapping generations), who base their predictions on genetic variation transmitted to descendants. This is because the approach using genetic variation cannot be sustained for general selection schemes. Santiago and Caballero (1995) suggest (their Equation 13) that a change in covariance between a general selective advantage and a neutral gene following selection is determined by the reduction in genetic variation. This is true for mass selection, where the index of selection is solely a function of the total breeding value and residual error, but will not be true in general (Woolliams *et al.* 1999). Bijma *et al.* (2000) show why there is agreement between the two approaches for mass selection in discrete generations and also why the current methods are required to cope with overlapping generations.

Prediction: Usable predictions were obtained by Wray *et al.* (1994) and an alternative form based upon Wray *et al.* (1994) was used by Villanueva and Woolliams (1997). However, the method of Wray *et al.* (1994) was complicated, although it attempted to model the expected proliferation of ancestral lines. The authors believe the proposed method is conceptually simpler than that of Wray *et al.* (1994) and is open to development.

In any attempt to obtain prediction formulas, a balance has to be achieved between accuracy and simplicity. We have used simple linear models to interpret the theory. Thus in application the prediction consists of two elements: (i) the squared expected contribution and (ii) the deviation from independent Poisson families. The first of these elements was applied precisely as described by Woolliams *et al.* (1999), with corrections for finite numbers only being used to obtain the sample variance of selective advantages. No other modifications were needed because the other terms in the squared

expected contribution were estimates of regression coefficients, which were assumed to be relatively robust to finite sampling. This assumption may be justified in part by the excellent agreement obtained by Woolliams *et al.* (1999) between simulations and deterministic predictions of expected long-term contributions. The second element, calculating the deviation from independent Poisson families, only required extension of the method of Burrows (1984) to two sexes. The correlation coefficients among full-sibs and half-sibs used for calculating this element were those obtained assuming infinite numbers but, to compensate for this, no reduction for finite samples was applied to the squared means.

The choice of selective advantages has as an objective the minimum number needed to make the selective processes in different time periods independent. Using sib indices as an example, the authors considered both the method presented, where only breeding values were included as selective advantages, and an alternative definition in which the selective advantages were the half-sib mean and deviation of the full-sib mean from the half-sib mean. The potential benefit from the alternative parameterization is that the environmental covariances in the index arising from the sib means are accounted for within the expected long-term contribution. Conditioning on the sib means is more than is strictly necessary for conditional independence between generations. However, while results using the alternative parameterization were as accurate in most cases (results not shown), the underestimates explored in the results tended to be more severe. One reason for this is that terms included in the expected long-term contribution are modeled by linear functions, whereas modeling the environmental correlations by the method of Burrows (1984) allows part of the nonlinearity to be accounted for. Therefore, the more terms that are included linearly in the expected long-term contribution, the greater the errors arising from nonlinearity.

Nonlinear relationships between the selective advantage and long-term contributions occurred when high selection intensities of selection were combined with moderate heritabilities, large numbers of parents, and high mating ratios. Results from including quadratic terms in the model for the expected long-term contribution (unpublished) confirm that the serious prediction errors arise from the assumption of linearity rather than from Equation 29.

There are good reasons to believe that these departures from linearity should not prove a major problem where the objective is to design effective breeding schemes. First, on pragmatic grounds the curvilinear relationship shown in Figure 1 suggests that 15% of selected males were being used with no expectation of long-term contribution to the population (this percentage is even higher if the contributions were plotted against the observed half-sib mean!). The resources used to keep and breed these animals are clearly wasted. In an ideal selection scheme, an ancestor's long-term contributions will be zero or, once its Mendelian sampling term is above a critical threshold, linearly related to the sampling term (Woolliams and Thompson 1994; Grundy *et al.* 1998). Consequently it would be expected that in an ideal scheme, the long-term contribution of a selected ancestor will show an approximate linearity with its breeding value. This argument suggests that if the design objective is for a scheme to generate gain efficiently from the resources available, a linear model for the relationship between the long-term contribution and the selective advantage should prove sufficient. If so, then the need for improved deterministic models to cater for the schemes with large prediction errors would be removed. The viewpoint that the schemes with large prediction errors are inefficient is supported by the results of Villanueva and Woolliams (1997), who showed that when using sib indices, efficient schemes had $d \leq 2$ for which the methods presented here had a good fit.

In conclusion, this article has (i) established a broader theorem (compared to Wray and Thompson 1990) concerning the relationship between squared long-term genetic contributions and rates of inbreeding, in particular extending the theorem to nonrandom mating and to overlapping generations; (ii) shown that, for random mating, the relationship can be generalized from long-term contributions that are simply observed to encompass expected long-term contributions that can be predicted; and (iii) shown how these equations might be interpreted with simple linear models in the context of predicting rates of inbreeding with sib indices in discrete generations. Together with the findings of Woolliams *et al.* (1999), the findings of this study show how rates of inbreeding may be predicted in general populations with complex structures and genetic models.

J.A.W. gratefully acknowledges financial support from the Ministry of Agriculture, Fisheries and Food (United Kingdom), and the support and encouragement of Prof. A. Maki-Tanila, who gave an opportunity for this work to be initiated. The contribution of P.B. was financially supported by the Netherlands Technology Foundation (STW) and coordinated by the Earth and Life Science Foundation (ALW).

LITERATURE CITED

Bijma, P., and J. A. Woolliams, 1999 Prediction of genetic contributions and generation intervals in populations with overlapping generations under selection. Genetics **151:** 1197–1210.

Bijma, P., J. A. M. van Arendonk and J. A. Woolliams, 2000 Prediction of rates of inbreeding in populations with discrete or overlapping generations under mass selection. Genetics **154:** 1865–1877.

Bulmer, M. G., 1980 *The Mathematical Theory of Quantitative Genetics.* Clarendon Press, Oxford.

Burrows, P. M., 1984 Inbreeding under selection from related families. Biometrics **40:** 895–906.

Caballero, A., and W. G. Hill, 1992 Effective size of nonrandom mating populations. Genetics **130:** 909–916.

Ghai, G. L., 1965 Structure of populations under mixed random and sib mating. Theor. Appl. Genet. **39:** 179–182.

Grundy, B., B. Villanueva and J. A. Woolliams, 1998 Dynamic

selection procedures for constrained inbreeding and their consequences for pedigree development. Genet. Res. **72:** 159–168.
Hill, W. G., 1972 Effective size of populations with overlapping generations. Theor. Popul. Biol. **3:** 278–289.
Hill, W. G., 1979 A note on effective population size with overlapping generations. Genetics **92:** 317–322.
Mendell, N. C., and R. C. Elston, 1974 Multifactorial traits: genetic analysis and prediction of recurrence risks. Biometrics **30:** 41–57.
Nomura, T., 1996 Effective size of selected populations with overlapping generations. J. Anim. Breed. Genet. **113:** 1–16.
Robertson, A., 1961 Inbreeding in artificial selection programmes. Genet. Res. **2:** 189–194.
Santiago, E., and A. Caballero, 1995 Effective size of populations under selection. Genetics **139:** 1013–1030.
Villanueva, B., and J. A. Woolliams, 1997 Optimization of breeding programmes under index selection and constrained inbreeding. Genet. Res. **69:** 145–158.
Villanueva, B., R. Pong-Wong, B. Grundy and J. A. Woolliams, 1999 Potential benefit from using an identified major gene in BLUP evaluation with truncation and optimal selection. Genet. Sel. Evol. **31:** 115–133.
Woolliams, J. A., and P. Bijma, 1999 Additional results and appendices to "Predicting rates of inbreeding in populations undergoing selection." http://www.ri.bbsrc.ac.uk/geneflow/ (29 July 1999).
Woolliams, J. A., and R. Thompson, 1994 A theory of genetic contributions. Proceedings of the 5th World Congress on Genetics Applied to Livestock Production, Vol. 19, Guelph, Canada, pp. 127–134.
Woolliams, J. A., N. R. Wray and R. Thompson, 1993 Prediction of long term contributions and inbreeding in populations undergoing mass selection. Genet. Res. **62:** 231–242.
Woolliams, J. A., P. Bijma and B. Villanueva, 1999 Expected genetic contributions and their impact on gene flow and genetic gain. Genetics **153:** 1009–1020.
Wray, N. R., and R. Thompson, 1990 Prediction of rates of inbreeding in selected populations. Genet. Res. **55:** 41–54.
Wray, N. R., J. A. Woolliams and R. Thompson, 1994 Prediction of rates of inbreeding in populations undergoing index selection. Theor. Appl. Genet. **87:** 878–892.

Communicating editor: R. G. Shaw

APPENDIX A: THE EXPECTED MENDELIAN SAMPLING VARIANCE

The expected Mendelian sampling variance in generation 1 summed over all alleles in the founders can be calculated using the following argument. For the progeny of the carrier founder i^* of the allele the gene frequency has mean $\frac{1}{4}$, *i.e.*, half of the gene frequency in carrier ($\frac{1}{2}$) plus half of that in mate (0), with $\sigma^2_a = \frac{1}{16}$. For progeny of other parents, $\sigma^2_a = 0$. Therefore, for a single allele, the Mendelian sampling variance is $\sigma^2_a = n_{i^*}/(16X)$, where n_{i^*} is the number of offspring of i^* selected in generation 1, and where X is the total number selected. Summing over all alleles (two per base individual), and since the sum of the number of offspring selected over all parents is $2X$, the expected variance is $\frac{1}{4}$.

At generation 2 and later, with true random mating the Mendelian sampling variance will be reduced. For dioecious species this will be delayed by a generation through nonrandom mating, and in general the expected variance is $\frac{1}{4}(1-\omega)(1-\Delta F)^{u-1}$ in generation $u > 1$, where ΔF is the rate of inbreeding among the parents.

APPENDIX B: EXTENSION OF THE PROOF RELATING EXPECTED CONTRIBUTIONS TO RATES OF INBREEDING TO INCLUDE OVERLAPPING GENERATIONS

Let X_q be the number of parents in category q and for convenience define a diagonal matrix $\boldsymbol{N}$ with elements X_q. The prediction for ΔF in overlapping generations is given by Equation 14. Let $\mu_{i(q)} = E[r_{i(q)}|\boldsymbol{s}_{i(q)}]$ and $\sigma^2_{i(q)} = \mathrm{Var}[r_{i(q)}|\boldsymbol{s}_{i(q)}]$ so that

$$\Delta F = \tfrac{1}{4}\sum_q X_q(E_s[\mu^2_{i(q)}] + E_s[\sigma^2_{i(q)}]). \tag{B1}$$

Let $\boldsymbol{m}$ be the vector with element q equal to $E_s[\mu^2_{i(q)}]$, $\boldsymbol{\nu}$ be the vector with element q equal to $E_s[\sigma^2_{i(q)}]$, and $\mathbf{1}$ be the vector with all elements equal to 1, so that $\Delta F = \frac{1}{4}\mathbf{1}^{\mathrm{T}}\boldsymbol{N}\boldsymbol{m} + \frac{1}{4}\mathbf{1}^{\mathrm{T}}\mathbf{N}\boldsymbol{\nu}$. In Equation 16, the n_i is no longer a single number but is a vector of numbers $\boldsymbol{n}_{i(q)}$, where the pth element $n_{p,i(q)}$ is the number of offspring of $i(q)$ selected that belong to category p.

Thus Equation 16 becomes

$$E[r_{i(q)}|\boldsymbol{s}_{i(q)},\boldsymbol{n}_{i(q)}] = \tfrac{1}{2}\sum_p n_{p,i(q)}E[r_{j(p)}|\boldsymbol{s}_i, j(p) \text{ offspring of } i(q)] \tag{B2}$$

$$\mathrm{Var}[r_{i(q)}|\boldsymbol{s}_{i(q)},\boldsymbol{n}_{i(q)}] = \tfrac{1}{4}\sum_p n_{p,i(q)}\mathrm{Var}[r_{j(p)}|\boldsymbol{s}_{i(q)}, j(p) \text{ offspring of } i(q)]. \tag{B3}$$

To obtain Equations 18 and 19, we need to define $\boldsymbol{\theta}_{n,i(q)} = E[\boldsymbol{n}_{i(q)}]$ with element p given by $\theta_{n,i(q),p}$, and $\boldsymbol{V}_{n,i(q)}$ to be the variance-covariance matrix for the elements $n_{i(q)}$, and to simplify the expressions define γ to be the vector with elements $\gamma_p = E[r_{j(p)}|\boldsymbol{s}_{i(q)}, j(p)$ offspring of $i(q)]$, and η to be a vector with elements $\eta_p = \mathrm{Var}[r_{j(p)}|\boldsymbol{s}_{i(q)}, j(p)$ offspring of $i(q)]$. This results in

$$\mu_{i(q)} = \tfrac{1}{2}\boldsymbol{\theta}^{\mathrm{T}}_{n,i(q)}\gamma \tag{B4}$$

$$\sigma^2_{i(q)} = \tfrac{1}{4}\boldsymbol{\theta}^{\mathrm{T}}_{n,i(q)}\boldsymbol{\eta} + \tfrac{1}{4}\gamma^{\mathrm{T}}\boldsymbol{V}_{n,i(q)}\gamma. \tag{B5}$$

It is now possible to make the same assumption as for discrete generations, *i.e.*, that the number in category p selected from parent $i(q)$ has a Poisson distribution:

$$\begin{aligned}\sigma^2_{i(q)} &= \tfrac{1}{4}\sum_p \theta_{n,i(q),p}(\eta_p + \gamma^2_p)\\ &= \tfrac{1}{4}\sum_p \theta_{n,i(q),p}E[r^2_{j(p)}|\boldsymbol{s}_{i(q)}, j(p) \text{ offspring of } i(q)].\end{aligned} \tag{B6}$$

Then by taking expectations over $\boldsymbol{s}_{i(q)}$ in (B6), and using the equilibrium property that $E_s[r^2_{i(q)}]$ is unchanged from generation to generation, show (Woolliams and Bijma 1999) that

$$\sum_{\text{categories } q} X_q E_s[\sigma^2_{i(q)}] = \tfrac{1}{2}\sum_{\text{categories } q} X_q E[r^2_{i(q)}|\boldsymbol{s}_{i(q)}] \tag{B7}$$

$$\sum_{\text{categories } q} X_q E_s[\sigma^2_{i(q)}] = \sum_{\text{categories } q} X_q E_s[\mu^2_{i(q)}]. \tag{B8}$$

Thus $\mathbf{1}^{\mathrm{T}}\boldsymbol{N}\boldsymbol{m} = \mathbf{1}^{\mathrm{T}}\boldsymbol{N}\boldsymbol{\nu}$ and substitution into Equation 17 gives the result

$$\Delta F = \tfrac{1}{2}\mathbf{1}^{\mathrm{T}}\boldsymbol{N}\boldsymbol{m} = \tfrac{1}{2}\sum_q X_q E_s[\mu^2_{(q)}]. \tag{B9}$$

APPENDIX C: PREDICTION OF EXPECTED GENETIC CONTRIBUTIONS FOR SIB INDICES

Expected genetic contributions were calculated using equilibrium genetic parameters. The genetic parameters were obtained by iterating rounds of selection starting from an unselected base generation with additive genetic variation h_0^2 and the phenotypic variance 1. The iterative equations were $\sigma^2_{A,t+1} = \sigma^2_{Am,t} + \sigma^2_{Af,t} + \frac{1}{2}h_0^2$ and $\sigma^2_{Ax,t} = \frac{1}{4}\sigma^2_{A,t}(1 - k_x\rho_t^2)$, where $\sigma^2_{A,t}$ is the additive genetic variance in generation t, ρ_t denotes the accuracy of selection in generation t (see Villanueva *et al.* 1999), and x = m or f as appropriate. Equilibrium variances were obtained in five iterations. In the remainder of this appendix the notation used is given in Table 2.

Calculation of the expected long-term genetic contributions followed the methods of Woolliams *et al.* (1999). Briefly these methods depend upon defining two regression models: the first describes the relative fitness of a parent as a linear function of its selective advantages; the second regression model describes the relationship of the selective advantages of the selected offspring with those of its parent. In discrete generations these models will depend only upon the sex of the parent and the sex of the selected offspring (in overlapping generations they may also depend on age).

For discrete generations the values of α_m and α_f are simply $(2X_m)^{-1}$ and $(2X_f)^{-1}$, respectively, and so the only term that needs more detailed description is the calculation of $\boldsymbol{\beta}$. $\boldsymbol{\beta}$ is a vector of three regression coefficients, the first (β_1) describing the regression of the long-term contribution of a selected male on its selective advantages $A_{i(hs)}$ and the remaining (β_2, β_3) describing the regression of the long-term contribution of a selected female on its two selective advantages ($A_{i(hs)}$, $A_{i(fs)}$). In the remainder of the appendix the selective advantages are indexed 1–3 as above.

$\boldsymbol{\beta}$ is derived from the formula of Woolliams *et al.* (1999), which has been simplified for application to discrete generations,

$$\boldsymbol{\beta} = \boldsymbol{N}^{-1}(\mathrm{I} - \tfrac{1}{2}\boldsymbol{\Pi}^{\mathrm{T}})(\tfrac{1}{2}\Lambda^{\mathrm{T}})(\tfrac{1}{2}, \tfrac{1}{2})^{\mathrm{T}},$$

where $\mathbf{N}$ is a diagonal matrix with elements (X_m, X_f, X_f), $\boldsymbol{I}$ is the identity matrix, and where $\boldsymbol{\Pi}$ and $\boldsymbol{\Lambda}$ are described below.

Λ is a (2 × 3) matrix, where λ_{i1} is the regression coefficient for the relative fitness of a male parent on its selective advantage, and where λ_{i2}, λ_{i3} are the corresponding coefficients for the selective advantages of a female parent. When $i = 1$ the relative fitness is for having male offspring selected and $i = 2$ for having female offspring selected. These coefficients will depend on the index of selection used and the selection intensity. The coefficients are derived using Appendix A of Woolliams *et al.* (1999). The elements are $\lambda_{11} = \lambda_{12} = i_m b_3 \sigma_I^{-1}$, $\lambda_{21} = \lambda_{22} = i_f b_3 \sigma_I^{-1}$, $\lambda_{13} = i_m b_2 \sigma_I^{-1}$, and $\lambda_{23} = i_f b_2 \sigma_I^{-1}$.

$\boldsymbol{\Pi}$ is a (3 × 3) matrix, with π_{ij} being the regression coefficient of selective advantage i of a selected offspring on the selective advantage j of the parent. This matrix describes exactly how the selection process in one generation is related to the same process in the next generation. The elements of $\boldsymbol{\Pi}$ are derived by standard selection theory (described in detail in Appendix B of Woolliams *et al.* 1999) and account for the effects of selection. Let $z = \rho_I\sigma_A/\sigma_I$; then the elements of $\boldsymbol{\Pi}$ are

$$\pi_{11} = \tfrac{1}{2}(1 - k_m b_3 z), \quad \pi_{12} = \tfrac{1}{2}(1 - k_m b_3 z), \quad \pi_{13} = \tfrac{1}{2}(1 - k_m b_2 z),$$
$$\pi_{21} = \tfrac{1}{2}(1 - k_f b_3 z)/d, \quad \pi_{22} = \tfrac{1}{2}(1 - k_f b_3 z)/d, \quad \pi_{23} = \tfrac{1}{2}(1 - k_f b_2 z)/d,$$
$$\pi_{31} = \tfrac{1}{2}(1 - k_f b_3 z)(1 - d^{-1}), \quad \pi_{32} = \tfrac{1}{2}(1 - k_f b_3 z)(1 - d^{-1}), \quad \pi_{33} = \tfrac{1}{2}(1 - k_f b_2 z)(1 - d^{-1}).$$

Example. For $X_m = 20$, $X_f = 60$, $h_0^2 = 0.4$, $n_m = n_f = 4$ with weights $b_1 = 1.0$, $b_2 = 1.5$, $b_3 = 2.0$. The principal parameters for the scheme are given in Table 3. Using the formulas given above,

$$\tfrac{1}{2}\boldsymbol{\Lambda} = (0.797, 0.797, 0.598|0.551, 0.551, 0.413)$$

$$\tfrac{1}{2}\boldsymbol{\Pi} = (0.123, 0.123, 0.155|0.045, 0.045,$$
$$0.055\ |0.090, 0.090, 0.109).$$

The solutions are $\boldsymbol{\alpha} = (0.0250, 0.0083)$ and $\boldsymbol{\beta} = (0.0447, 0.0149, 0.0130)$.

APPENDIX D: THE VARIANCES OF FAMILY SIZE AFTER SELECTION WHEN LITTER SIZES ARE CONSTANT

The variances of family size when litter sizes are constant are derived by combining results of Burrows (1984) and Mendell and Elston (1974), which extend and formalize results used by Woolliams *et al.* (1993). For simplicity, litters are assumed to have n males and n females, and there are T candidates for selection in each sex. The basic approach of using factorial moments, *i.e.*, $E[n_{ij}(q)(n_{ij}(q) - 1)]$, where $n_{ij}(q)$ is the number of sex q (*i.e.*, q = m or f) selected from the full-sib family with sire i and dam j, was described in detail by Burrows (1984). Since Burrows (1984) was working in the context of forestry only a single sex was considered and hence some extension to two sexes is necessary. The approach of Burrows (1984) has been preferred since it results in elegant formulas.

Denote $n_{ij}(q)$ as the number of offspring selected of sex q from the full-sib family of sire i and dam j, and $n_{i*}(q)$ as the number selected from sire i (*i.e.*, summed over all its mates). Note that the variance of family size can be simply expressed in terms of the factorial moments:

$$\mathrm{Var}[n_{ij}(q)] = E[n_{ij}(q)(n_{ij}(q) - 1)] - E[n_{ij}(q)](E[n_{ij}(q)] - 1) \quad \text{(D1)}$$

$$\mathrm{Var}[n_{i*}(q)] = E[n_{i*}(q)(n_{i*}(q) - 1)] - E[n_{i*}(q)](E[n_{i*}(q)] - 1). \quad \text{(D2)}$$

To obtain deviations of the variance from Poisson family

size, the term in $E[n_{ij}(q)](E[n_{ij}(q)] - 1)$ in Equation D1 is replaced by $E[n_{ij}(q)]^2$ and a similar change is made in Equation D2.

Burrows (1984) derived the asymptotic form (Burrows 1984, Equations 4–12),

$$E[n_{ij}(q)(n_{ij}(q) - 1)] = [n(n-1)X_q(X_q - 1)]/[T(T-1)R(p_q, \rho_{FS})], \quad \text{(D3)}$$

where X_q is the total number of that sex selected and T is the total number of candidates, p_q the proportion selected (*i.e.*, X_q/T), and ρ_{FS} the correlation between full-sibs. $R(p, \rho)$ is the ratio $p^2/\Phi(v, v; \rho)$, where $\Phi(v, v; \rho)$ is the upper-quadrant probability that both variables of a standardized bivariate normal distribution with correlation coefficient ρ exceed v; v is defined by $\Phi(v) = 1 - p$; and $\Phi(v)$ is the distribution function of the standard univariate normal distribution. The ratio is essentially the probability of two sibs being above the index truncation point when the index correlation among sibs is zero (as in random selection) divided by the probability with correlation ρ. Burrows (1984) uses tabulated values for the coselection ratio $R(p, \rho)$, but these values can be approximated closely by results from Mendell and Elston (1974): $R(p, \rho) = p/\Phi[(i\rho - v)(1 - k\rho^2)^{-1/2}]$. This approximation is used throughout. To allow extension to two sexes we denote $R(p, \rho)$ by $R(p, p, \rho)$. [It seems more natural to the authors to use a term equal to $1/R(p, p, \rho)$ in the formula, which describes the proportional increase in coselection; however, we have used $R(p, p, \rho)$ to maintain continuity of notation with Burrows (1984).]

Burrows (1984) derived the additional result to use for the variance of half-sib family sizes. In this article only paternal half-sib families are considered,

$$E[n_{ij}(q)n_{ik}(q)] = [n^2X_q(X_q - 1)]/[T(T-1)R(p_q, p_q, \rho_{HS})], \quad \text{(D4)}$$

where j and k are distinct mates to a common parent i, and ρ_{HS} is the correlation between half-sibs. Therefore, for a sire with d mates,

$$\begin{aligned} E[n_{i*}(q)(n_{i*}(q) - 1)] &= \sum_j E[n_{ij}(q)(n_{ij}(q) - 1)] + \sum_{j \neq k} E[n_{ij}(q)n_{ik}(q)] \\ &= [d(1 - n^{-1})/R(p_q, p_q, \rho_{FS}) + d(d-1)/R(p_q, p_q, \rho_{HS})] \times [n^2X_q(X_q - 1)]/[T(T-1)]. \end{aligned} \quad \text{(D5)}$$

The covariances of male and female family size are

$$\mathrm{Cov}[n_{ij}(m), n_{ij}(f)] = E[n_{ij}(m)n_{ij}(f)] - E[n_{ij}(m)]E[n_{ij}(f)] \quad \text{(D6)}$$

$$\mathrm{Cov}[n_{i*}(m), n_{i*}(f)] = E[n_{i*}(m)n_{i*}(f)] - E[n_{i*}(m)]E[n_{i*}(f)]. \quad \text{(D7)}$$

The expected cross-products are derived analogously to the variance and are given by

$$E[n_{ij}(m)n_{ij}(f)] = [n^2X_mX_f]/[T^2R(p_m, p_f, \rho_{FS})] \quad \text{(D8)}$$

$$E[n_{i*}(m)n_{i*}(f)] = [d/R(p_m, p_f, \rho_{FS}) + d(d-1)/R(p_m, p_f, \rho_{HS})] \times [n^2X_mX_f/T^2]. \quad \text{(D9)}$$

The rationale of the term $R(p_m, p_f, \rho)$ as a ratio of probabilities for random selection and with correlation ρ remains unchanged but has been extended to the situation with two sexes with different selection proportions. This ratio is calculated from Mendell and Elston (1974) using $R(p_m, p_f, \rho) \approx p_f/\Phi[(i_m\rho - v_f)(1 - k_m\rho^2)^{-1/2}]$, which was found by Wray *et al.* (1994) to be the more accurate implementation of their results (there are two possible implementations since the approximation is asymmetric in male and female parameters).

To obtain the variances and covariances conditional upon the selective advantage, the regression model derived for the expected number of offspring selected is used (see appendix c).

Thus, for a dam family,

$$E[n_{ij}(m)]^2 = d^{-2}(1 + \lambda_{12}^2 v(A_{i(hs)}) + \lambda_{13}^2 v(A_{i(fs)})) \quad \text{(D10)}$$

$$E[n_{ij}(m)]E[n_{ij}(f)] = d^{-1}(1 + \lambda_{12}\lambda_{22} v(A_{i(hs)}) + \lambda_{13}\lambda_{23} v(A_{i(fs)})) \quad \text{(D11)}$$

$$E[n_{ij}(f)]^2 = 1 + \lambda_{22}^2 v(A_{i(hs)}) + \lambda_{23}^2 v(A_{i(fs)}), \quad \text{(D12)}$$

and for a sire family,

$$E[n_{i*}(m)]^2 = 1 + \lambda_{11}^2 v(A_{i(hs)}) \quad \text{(D13)}$$

$$E[n_{i*}(m)]E[n_{i*}(f)] = d(1 + \lambda_{11}\lambda_{21} v(A_{i(hs)})) \quad \text{(D14)}$$

$$E[n_{i*}(f)]^2 = d^2(1 + \lambda_{21}^2 v(A_{i(hs)})). \quad \text{(D15)}$$

Example. For $X_m = 20$, $X_f = 60$, $h_0^2 = 0.4$, $n_m = n_f = 4$ with weights $b_1 = 1.0$, $b_2 = 1.5$, $b_3 = 2.0$, most parameters for this scheme are given in Table 3, and the λ values are derived in appendix c. The coselection ratios are $R(p_m, p_m, \rho_{FS}) = 0.350$, $R(p_m, p_m, \rho_{HS}) = 0.546$, $R(p_m, p_f, \rho_{FS}) = 0.482$, $R(p_m, p_f, \rho_{HS}) = 0.656$, $R(p_f, p_f, \rho_{FS}) = 0.589$, $R(p_f, p_f, \rho_{HS}) = 0.742$.

Applying the results of this appendix gives $\mathbf{V}_{n(m),\mathrm{dev}} = (0.186,\ 0.751|0.751,\ -0.079)$ and $\mathbf{V}_{n(f),\mathrm{dev}} = (0.020,\ 0.159|0.159,\ -0.154)$.